

Towards Optimal Measurement of Power Spectra II: A Basis of Positive, Compact, Statistically Orthogonal Kernels

A. J. S. Hamilton

JILA and Dept. of Astrophysical, Planetary and Atmospheric Sciences, Box 440, University of Colorado, Boulder CO 80309, USA
email: *ajsh@dark.colorado.edu*

1 February 2008

ABSTRACT

This is the second of two papers which address the problem of measuring the unredshifted power spectrum of fluctuations from a galaxy survey in optimal fashion. A key quantity is the Fisher matrix, which is the inverse of the covariance matrix of minimum variance estimators of the power spectrum of the survey. It is shown that bases of kernels which give rise to complete sets of statistically orthogonal windowed power spectra are obtained in general from the eigenfunctions of the Fisher matrix scaled by some arbitrary positive definite scaling matrix. Among the many possible bases of kernels, there is a basis, obtained by applying an infinitely steep scaling function, which leads to kernels which are positive and compact in Fourier space. This basis of kernels, along with the associated minimum variance pair weighting derived in the previous paper, would appear to offer a solution to the problem of how to measure the unredshifted power spectrum optimally. Illustrative kernels are presented for the case of the PSCz survey.

Key words: cosmology: theory – large-scale structure of Universe

1 INTRODUCTION

This is the second of a pair of papers (Hamilton 1997, hereafter Paper 1) concerned with the problem of measuring the unredshifted power spectrum of fluctuations in a galaxy survey in optimal fashion.

As discussed by Tegmark, Taylor & Heavens (1997), a fundamental quantity in the estimation of any set of parameters from a data set is the Fisher information matrix, which is the matrix of second derivatives of minus the log-likelihood function with respect to the parameters, and whose inverse gives an estimate of the covariance matrix of the parameters. It was demonstrated in Paper 1 that if the parameter set of interest is the unredshifted power spectrum of fluctuations in a galaxy survey, and under the standard assumption that the main source of uncertainty in the power spectrum is cosmic variance plus Poisson sampling noise, then the Fisher matrix can be evaluated rather directly, albeit in an approximation. With the Fisher matrix in hand, it becomes possible to construct complete sets of kernels which yield statistically orthogonal sets of windowed power spectra. Such bases of kernels hold the potential of providing an optimal way to measure the power spectrum, in the sense that they offer statistical orthogonality (which implies statistical independence in the case of Gaussian fluctuations), and complete-

ness (hence maximum resolution, since the power spectrum cannot be decomposed further into statistically orthogonal subunits). The purpose of the present paper is to explore this idea further, and in particular to show some actual examples. I choose to use the geometry and selection function of PSCz survey (W. Saunders et al., in preparation) for illustration, and I assume Gaussian fluctuations for simplicity.

As shown in §2, there are in fact many distinct bases of kernels which give rise to statistically orthogonal sets of windowed power spectra. After giving details of the computation of the Fisher matrix in §3, I show two trial bases of kernels in §4. It is found there that, without special attention, kernels are liable to be broad and wiggly in Fourier space, which are not desirable properties. The difficulty is resolved in §5, where it is shown how to construct a basis of kernels which are compact and everywhere positive in Fourier space. Section 5 leaves unanswered a number of basic questions as to the generality and uniqueness of the proposed construction. These questions are posed §6. The results are summarized in §7.

2 STATISTICALLY ORTHOGONAL SETS OF POWER SPECTRA

Let ξ_α denote the power spectrum $\xi(k)$ windowed through a set of kernels $\psi_\alpha(k)$:

$$\xi_\alpha \equiv \int_0^\infty \psi_\alpha(k) \xi(k) 4\pi k^2 dk / (2\pi)^3. \quad (1)$$

The power spectrum $\xi(k)$ is real, and the kernels $\psi_\alpha(k)$ can also be taken to be real without loss of generality. Let $\hat{\xi}_\alpha$ denote the minimum variance estimate of the windowed power spectrum ξ_α (as in Paper 1, the hat on $\hat{\xi}$ is used to distinguish the estimator from the true power spectrum ξ)

$$\begin{aligned} \hat{\xi}_\alpha &\approx \hat{\xi}_\alpha, \\ \hat{\xi}_\alpha &\equiv \int W(\mathbf{r}_i, \mathbf{r}_j) \delta(\mathbf{r}_i) \delta(\mathbf{r}_j) d^3r_i d^3r_j \end{aligned} \quad (2)$$

where $W(\mathbf{r}_i, \mathbf{r}_j)$ is the minimum variance pair window derived in Paper 1, equation (39), for the power spectrum windowed through the kernel $G(k) = \psi_\alpha(k)$. The integration in equation (2) is over all pairs of observed overdensities $\delta(\mathbf{r}_i) \delta(\mathbf{r}_j)$ at positions \mathbf{r}_i and \mathbf{r}_j in the survey. The estimators $\hat{\xi}_\alpha$ of the windowed power spectra ξ_α are said to be statistically orthogonal if the expectation value of the covariance between different elements vanishes,

$$\langle \Delta \hat{\xi}_\alpha \Delta \hat{\xi}_\beta \rangle = 0 \quad (\alpha \neq \beta). \quad (3)$$

Statistical orthogonality does not imply statistical independence in general, but it does imply statistical independence for the particular case of Gaussian fluctuations, for which by definition all the higher order irreducible moments vanish.

There are not one but many distinct bases, i.e. complete linearly independent sets, of kernels ψ_α which define statistically orthogonal sets of power spectra $\hat{\xi}_\alpha$. To see this, consider first that the eigenfunctions of the expected covariance matrix $\langle \Delta \hat{\xi}(k) \Delta \hat{\xi}(k') \rangle$ of estimators $\hat{\xi}(k)$ of the power spectrum define a certain complete orthonormal set of kernels $\psi_\alpha(k)$, for which by construction the windowed power spectra $\hat{\xi}_\alpha$ are statistically orthogonal. Now rescale each kernel $\psi_\alpha(k)$ to unit variance, by dividing it by the square root of the variance, $\psi_\alpha \rightarrow \psi_\alpha / \langle \Delta \hat{\xi}_\alpha^2 \rangle^{1/2}$. The covariance matrix of the power spectrum windowed through the resulting basis of scaled kernels is the unit matrix. Any orthogonal rotation amongst these scaled kernels yields another basis of kernels, which defines another distinct statistically orthogonal set of windowed power spectra.

If $\psi_\alpha(k)$ is a set of kernels which defines a statistically orthogonal set of estimators $\hat{\xi}_\alpha$ of windowed power spectra, then evidently $c_\alpha \psi_\alpha(k)$, where c_α is any set of finite nonzero normalization constants, defines a renormalized version of essentially the same statistically orthogonal set, with $\hat{\xi}_\alpha \rightarrow c_\alpha \hat{\xi}_\alpha$ and $\langle \Delta \hat{\xi}_\alpha^2 \rangle \rightarrow c_\alpha^2 \langle \Delta \hat{\xi}_\alpha^2 \rangle$. In the development below it is informative to carry along a set of normalization constants c_α . The normalization constants can be chosen arbitrarily, so that one is free to normalize the kernels $\psi_\alpha(k)$ in whatever manner seems convenient. In the present paper I choose to plot the kernels quadratically normalized, so that their squares have unit volume in k -space,

$$\int_0^\infty \psi_\alpha(k)^2 4\pi k^2 dk / (2\pi)^3 = 1. \quad (4)$$

In practical applications, notably when using the basis of positive kernels presented in §5, one might prefer to normalize the kernels linearly to unit volume in k -space, $\int_0^\infty \psi_\alpha(k) 4\pi k^2 dk / (2\pi)^3 = 1$. The difficulty with the linear normalization – more a matter of plotting rather than anything fundamental – is that for kernels $\psi_\alpha(k)$ which are not everywhere positive in k -space, such as those considered in §4, a linear rather than quadratic normalization tends to make noisy kernels wiggle more fiercely, obscuring the better-behaved kernels.

In a moment I will prove the following theorem. For any complete set of kernels $\psi_\alpha(k)$, along with any set of normalization constants c_α , yielding a statistically orthogonal set of windowed power spectra $\hat{\xi}_\alpha$, there exists a strictly positive definite – i.e., all the eigenvalues are strictly positive – real symmetric scaling matrix $\gamma(k, k')$ such that the scaled kernels $\phi_\alpha(k)$ defined by

$$\phi_\alpha(k) \equiv \int_0^\infty \gamma(k, k') c_\alpha \psi_\alpha(k') 4\pi k'^2 dk' / (2\pi)^3 \quad (5)$$

form a complete orthonormal basis which diagonalizes the scaled covariance matrix $\gamma^{-1} \langle \Delta \hat{\xi} \Delta \hat{\xi} \rangle \gamma^{-1}$.

The converse of this theorem is evident. If γ is any strictly positive definite real symmetric matrix, then the scaled covariance matrix $\gamma^{-1} \langle \Delta \hat{\xi} \Delta \hat{\xi} \rangle \gamma^{-1}$ defines a complete orthonormal set of eigenfunctions ϕ_α , with eigenvalues λ_α :

$$\gamma^{-1} \langle \Delta \hat{\xi} \Delta \hat{\xi} \rangle \gamma^{-1} \phi_\alpha = \lambda_\alpha \phi_\alpha. \quad (6)$$

The kernels ψ_α defined by $\psi_\alpha = c_\alpha^{-1} \gamma^{-1} \phi_\alpha$, the inverse of equation (5), then yield estimators $\hat{\xi}_\alpha$ of windowed power spectra, equations (1) and (2), which are statistically orthogonal because

$$\begin{aligned} \langle \Delta \hat{\xi}_\alpha \Delta \hat{\xi}_\beta \rangle &\equiv \psi_\alpha \langle \Delta \hat{\xi} \Delta \hat{\xi} \rangle \psi_\beta = c_\alpha^{-1} \phi_\alpha \gamma^{-1} \langle \Delta \hat{\xi} \Delta \hat{\xi} \rangle \gamma^{-1} \phi_\beta c_\beta^{-1} \\ &= c_\alpha^{-2} \lambda_\alpha \delta_{D\alpha\beta} \end{aligned} \quad (7)$$

is a diagonal matrix (no summation over α or β implied). Here $\delta_{D\alpha\beta}$ denotes the unit matrix. In particular, the variances $\langle \Delta \hat{\xi}_\alpha^2 \rangle$ of the windowed power spectra $\hat{\xi}_\alpha$ are related to the eigenvalues λ_α of the scaled covariance matrix $\gamma^{-1} \langle \Delta \hat{\xi} \Delta \hat{\xi} \rangle \gamma^{-1}$ by

$$\langle \Delta \hat{\xi}_\alpha^2 \rangle = \frac{\lambda_\alpha}{c_\alpha^2}. \quad (8)$$

As seen in Paper 1, what one actually calculates for a survey is not the expected covariance matrix $\langle \Delta \hat{\xi} \Delta \hat{\xi} \rangle$ of estimates of the power spectrum, but rather its inverse the Fisher matrix $T = \langle \Delta \hat{\xi} \Delta \hat{\xi} \rangle^{-1}$. The inverse of the scaled covariance matrix $\gamma^{-1} \langle \Delta \hat{\xi} \Delta \hat{\xi} \rangle \gamma^{-1}$ is the scaled Fisher matrix $\gamma T \gamma$. Thus the eigenfunctions ϕ_α of the scaled covariance matrix can in practice be constructed as the orthonormal eigenfunctions of the scaled Fisher matrix,

$$\gamma T \gamma \phi_\alpha = \mu_\alpha \phi_\alpha, \quad (9)$$

with eigenvalues $\mu_\alpha = 1/\lambda_\alpha$ which are the reciprocals of the eigenvalues of the scaled covariance matrix. The kernels ψ_α follow as before from $\psi_\alpha = c_\alpha^{-1} \gamma^{-1} \phi_\alpha$, the inverse of equation (5). The variances $\langle \Delta \hat{\xi}_\alpha^2 \rangle$ of the windowed power spectra $\hat{\xi}_\alpha$ are related to the eigenvalues μ_α of the scaled Fisher matrix by, equation (8),

$$\langle \Delta \hat{\xi}_\alpha^2 \rangle = \frac{1}{c_\alpha^2 \mu_\alpha}. \quad (10)$$

Thus the general procedure for constructing complete bases of kernels yielding statistically orthogonal sets of power spectra is as follows: Choose any strictly positive definite symmetric scaling matrix γ ; find the eigenfunctions ϕ_α of the scaled Fisher matrix $\gamma T \gamma$; and construct the kernels $\psi_\alpha \propto \gamma^{-1} \phi_\alpha$, normalizing to taste.

The next few paragraphs prove the previously stated theorem, which asserts that a scaling matrix γ exists for any complete set of kernels yielding a statistically orthogonal set of power spectra, and which therefore guarantees that the procedure of the previous paragraph is in fact general. Suppose then that $\psi_\alpha(k)$ comprise a complete set of kernels which give rise to a statistically orthogonal set $\hat{\xi}_\alpha$ of windowed power spectra, equations (1), (2), and (3). The kernels $\psi_\alpha(k)$ themselves need not be orthogonal, but they must be linearly independent. For if one of the kernels depended linearly on the others, say $\psi_\alpha = \sum_{\beta \neq \alpha} a_\beta \psi_\beta$ for some a_β , then the corresponding windowed power spectrum would likewise depend linearly on the others, $\hat{\xi}_\alpha = \sum_{\beta \neq \alpha} a_\beta \hat{\xi}_\beta$, so that $\langle \Delta \hat{\xi}_\alpha \Delta \hat{\xi}_\beta \rangle = a_\beta \langle \Delta \hat{\xi}_\beta^2 \rangle$ would be nonzero for some β , contradicting the statistical orthogonality of the $\hat{\xi}_\alpha$. Linear independence, coupled with the assumption of completeness, means that the kernels $\psi_\alpha(k)$ form a basis.

Define the real symmetric matrix $A(k, k')$ by

$$A(k, k') \equiv \sum_\alpha c_\alpha^2 \psi_\alpha(k) \psi_\alpha(k') . \quad (11)$$

Being real and symmetric, the matrix A is diagonalized by some orthogonal transformation O , so that $O^T A O$ is diagonal. Rotating equation (11) by the orthogonal transformation O , one concludes that the matrix A must be positive definite – that is, all its eigenvalues are positive or zero – since each eigenvalue is the inner product of a vector $O c \psi$ with itself. The assumption that the kernels ψ_α form a complete set implies further that the eigenvalues of A are all strictly positive. For equations (1) and (11) together imply

$$\int_0^\infty A(k, k') \xi(k') 4\pi k'^2 dk' / (2\pi)^3 = \sum_\alpha c_\alpha^2 \psi_\alpha(k) \xi_\alpha , \quad (12)$$

and if there existed a nontrivial eigenfunction $\xi(k)$ of A with zero eigenvalue, then the left hand side of equation (12) would vanish for this eigenfunction. The consequent vanishing of the right hand side of equation (12) would then imply either that the kernels $\psi_\alpha(k)$ are linearly dependent, which is false as shown above, or else that $\xi_\alpha = 0$ for all α for the eigenfunction $\xi(k)$. But if, as postulated, the kernels $\psi_\alpha(k)$ form a complete set, then $\xi(k) = \sum_\alpha x_\alpha \psi_\alpha(k)$ for some set of coefficients x_α , so that $\int \xi(k)^2 4\pi k^2 dk / (2\pi)^3 = \sum_\alpha x_\alpha \int \psi_\alpha(k) \xi(k) 4\pi k^2 dk / (2\pi)^3 = \sum_\alpha x_\alpha \xi_\alpha = 0$, whose only possible solution is $\xi(k) = 0$. Thus the matrix A can have no nontrivial eigenfunctions with zero eigenvalue, and hence A must be strictly positive definite, as claimed.

It follows that the matrix A of equation (11) can be written in the form $A = O \Lambda^2 O^T = (O \Lambda O^T)^2 = \gamma^{-2}$ where Λ is a diagonal matrix, and $\gamma \equiv O \Lambda^{-1} O^T$. Without loss of generality, all the square roots in $\Lambda = (\Lambda^2)^{1/2}$ can be taken to be positive. Thus the matrix γ , which I call the scaling matrix, is also a strictly positive definite real symmetric matrix. It then follows from equation (11) that the scaled kernels $\phi_\alpha(k)$ defined by equation (5) form a complete orthonormal set

$$\sum_\alpha \phi_\alpha(k) \phi_\alpha(k') = (2\pi)^3 \delta_D(k - k') \quad (13)$$

the Dirac delta-function $(2\pi)^3 \delta_D(k - k')$ being the unit matrix in the Fourier representation. In accordance with the convention of Paper 1, the Dirac delta-function in equation (13) is normalized on a 3-dimensional scale, so that $\int \delta_D(k - k') 4\pi k'^2 dk' = 1$. Equation (13) shows that $\phi_\alpha(k)$ is in effect an orthogonal matrix $\phi_{k\alpha}$, whose inverse is its transpose. Taking the factors in opposite order yields the equivalent orthonormality condition

$$\int_0^\infty \phi_\alpha(k) \phi_\beta(k) 4\pi k^2 dk / (2\pi)^3 = \delta_{D\alpha\beta} \quad (14)$$

where $\delta_{D\alpha\beta}$ is the unit matrix in the ϕ -representation.

The desired scaling matrix γ has thus been constructed, as asserted by the theorem, with the property that the scaled kernels ϕ_α defined by equation (5) form a complete orthonormal set. The final assertion of the theorem, that the scaled covariance matrix $\gamma^{-1} \langle \Delta \hat{\xi} \Delta \hat{\xi} \rangle \gamma^{-1}$ is diagonal with respect to the scaled kernels ϕ_α , follows directly, equation (7), from the hypothesis that the covariance matrix $\langle \Delta \hat{\xi} \Delta \hat{\xi} \rangle$ is diagonal with respect to the kernels ψ_α . This completes the proof of the theorem.

From the definition (1) of ξ_α , and the orthonormality condition (13) on the scaled kernels $\phi_\alpha(k)$ defined by equation (5), it follows that the power spectrum $\xi(k)$ can be expanded in terms of the kernels $\psi_\alpha(k)$ as

$$\xi(k) = \sum_\alpha \xi_\alpha c_\alpha^2 \int_0^\infty \gamma^2(k, k') \psi_\alpha(k') 4\pi k'^2 dk' / (2\pi)^3 \quad (15)$$

[note $\gamma^2(k, k')$ here is not $[\gamma(k, k')]^2$ but rather $\gamma^2(k, k') = \int_0^\infty \gamma(k, k'') \gamma(k'', k') 4\pi k''^2 dk'' / (2\pi)^3$]. The corresponding minimum variances estimators $\hat{\xi}(k)$ and $\hat{\xi}_\alpha$ are similarly related, by an expression which is essentially identical to (15).

Being real and symmetric, the scaling matrix $\gamma(k, k')$ must be diagonal in some representation or other. All the scaling matrices considered in the present paper are diagonal in Fourier space, so that the scaling matrix becomes a scaling function $\gamma(k)$. In that case, the relation (5) between the eigenfunctions $\phi_\alpha(k)$ and the kernels $\psi_\alpha(k)$ simplifies to

$$\phi_\alpha(k) = \gamma(k) c_\alpha \psi_\alpha(k) , \quad (16)$$

the orthogonality condition (14) becomes

$$\int_0^\infty \psi_\alpha(k) \psi_\beta(k) \gamma(k)^2 \frac{4\pi k^2 dk}{(2\pi)^3} = \begin{cases} 0 & (\alpha \neq \beta) \\ c_\alpha^{-2} & (\alpha = \beta) \end{cases} , \quad (17)$$

and the expansion (15) of the power spectrum $\xi(k)$ in the kernels $\psi_\alpha(k)$ reduces to

$$\xi(k) = \sum_\alpha \xi_\alpha c_\alpha^2 \psi_\alpha(k) \gamma(k)^2 . \quad (18)$$

The eigenfunctions $\phi_\alpha(k)$ of the scaled Fisher matrix $\gamma T \gamma$ are found by solving the eigenvalue equation (9), which reduces to

$$\int_0^\infty \gamma(k) T(k, k') \gamma(k') \phi_\alpha(k') \frac{4\pi k' dk'}{(2\pi)^3} = \mu_\alpha \phi_\alpha(k) . \quad (19)$$

The relation (10) between the variances $\langle \Delta \hat{\xi}_\alpha^2 \rangle$ and the eigenvalues μ_α of the scaled Fisher matrix remains unchanged.

3 COMPUTATION OF THE FISHER MATRIX

A perturbation procedure for computing the Fisher matrix $T(k, k')$ of the unredshifted power spectrum of a galaxy survey has been described in Paper 1. Here I assume for simplicity Gaussian fluctuations, and the zeroth order approximation, which is essentially the classical approximation considered by Feldman, Kaiser & Peacock (1994, hereafter FKP), where position and wavelength are simultaneously measurable.

To illustrate the procedures proposed in this paper, I take the geometry of the PSCz survey, which is the redshift survey of IRAS galaxies complete to 0.6 Jy (Saunders et al., in preparation). I take the angular geometry of PSCz to be the same as that of the QDOT survey (Lawrence et al. 1997), which covers most of the sky above galactic latitude $|b| > 10^\circ$. I take the radial selection function to be 6 times that of QDOT, which was designed to be a 1-in-6 random sampling of IRAS galaxies to 0.6 Jy. The QDOT radial selection function is measured using Turner's (1979) method, after making cosmological corrections for the volume element (assuming a flat Universe), and for evolution of the luminosity function (adopting luminosity evolution $L \propto (1+z)^{1.8}$). I chose also to exclude the local region closer than $25 h^{-1}\text{Mpc}$, to eliminate possible bias from the local overdensity; in practice this has little effect at the long wavelengths illustrated in this paper.

One test of the accuracy of the zeroth order approximation to the Fisher matrix adopted here is that the Fisher matrix should be strictly positive definite, that is, all its eigenvalues should be strictly positive. In the PSCz survey, I find that if the approximate Fisher matrix is computed with a resolution in Fourier space of $\Delta k = \pi/(896 h^{-1}\text{Mpc}) = 0.00351 h\text{Mpc}^{-1}$, or coarser, then all the eigenvalues are positive. At finer resolutions an increasing fraction of the eigenvalues are negative. For the trial bases exhibited in §4, I chose to over-resolve the kernels, using $\Delta k = \pi/(2048 h^{-1}\text{Mpc}) = 0.00153 h\text{Mpc}^{-1}$, which appears to do no harm to kernels of low variance. At this resolution $\sim 22\%$ of the eigenvalues of the Fisher matrix are negative, but the absolute values of the negative eigenvalues are all less than 10^{-4} of the largest eigenvalue.

The zeroth order approximation to the Fisher matrix is given in Paper 1, equation (86), as

$$T(k, k') = \frac{1}{2} \int |U(\mathbf{k} + \mathbf{k}')|^2 \text{d}o' / (4\pi)^2 \quad (20)$$

where $\text{d}o$ denotes an interval of solid angle and the integration is over all directions of \mathbf{k} and \mathbf{k}' , and where $U(\mathbf{k}) = \int U(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}} \text{d}^3r$ is the Fourier transform of the FKP window (Paper 1, eq. [72])

$$U(\mathbf{r}) = \frac{\Phi(\mathbf{r})}{1 + \xi(k_0)\Phi(\mathbf{r})}. \quad (21)$$

Here $\Phi(\mathbf{r})$ is the selection function, the expected number density of galaxies at position \mathbf{r} in the survey, and $\xi(k_0)$ is the assumed prior power spectrum at wavenumber k_0 . The higher order terms of the Fisher matrix $T(k, k')$ comprise an expansion in $\xi(k) - \xi(k_0)$ (Paper 1). Thus the greatest accuracy in the zeroth order term is accomplished by choosing k_0 close to k and k' ; I adopt

$$k_0 = (k + k')/2. \quad (22)$$

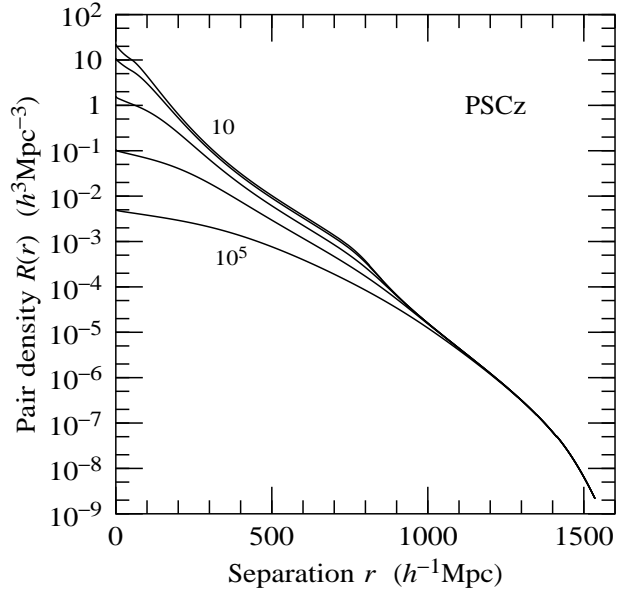


Figure 1. Expected density $R(r)$ of weighted random pairs at separation r , equation (24), in the PSCz survey. Curves, from top to bottom, are for prior powers $\xi(k_0) = 10, 10^2, 10^3, 10^4$, and $10^5 h^{-3}\text{Mpc}^3$ in the survey window $U(\mathbf{r})$, equation (21).

The approximation should be quite good, since the Fisher matrix (20) is almost diagonal in Fourier space, its largest elements satisfying $k \approx k'$.

The power spectrum $\xi(k_0)$ in the survey window (21) can be regarded as a prior in Bayesian model testing, or alternatively as a reasonable estimate which may in practice lead to kernels yielding nearly statistically orthogonal sets of windowed power spectra. I adopt the unredshifted power spectrum proposed by Peacock (1997). Peacock gives analytic formulae for the power spectrum in two cases, a flat model with $\Omega = 1$, and an open model with $\Omega = 0.3$; the results shown in the present paper are for the flat model.

It is convenient to evaluate the Fisher matrix (20) via an excursion back into real space. In the real representation, the Fisher matrix (20) is

$$T(r, r') = \frac{1}{2} \delta_D(r - r') R(r) \quad (23)$$

where $R(r)$ is the integral of the pair window $U(\mathbf{r}_i)U(\mathbf{r}_j)$ over all pairs ij in the survey separated by r

$$R(r) = \int U(\mathbf{r}_i)U(\mathbf{r}_j) \delta_D(|\mathbf{r}_i - \mathbf{r}_j| - r) \text{d}^3r_i \text{d}^3r_j. \quad (24)$$

Beware of equation (23)! It does not signify that the Fisher matrix is diagonal in real space, because the prior power spectrum $\xi(k_0)$ in the survey window $U(\mathbf{r})$, equation (21), is a different constant for different elements of the Fisher matrix $T(k, k')$ (the Fisher matrix would be diagonal in real space if the prior power spectrum were that of shot noise, $\xi(k) = \text{constant}$, but that is not the case here). In accordance with the convention of Paper 1, the Dirac delta-functions in equations (23) and (24) are normalized on a 3-dimensional scale, so that $\int \delta_D(r - r') 4\pi r'^2 \text{d}r' = 1$. The pair density $R(r)$, which in the literature is often denoted $\langle RR \rangle$ (for random-random), is the expected number of weighted

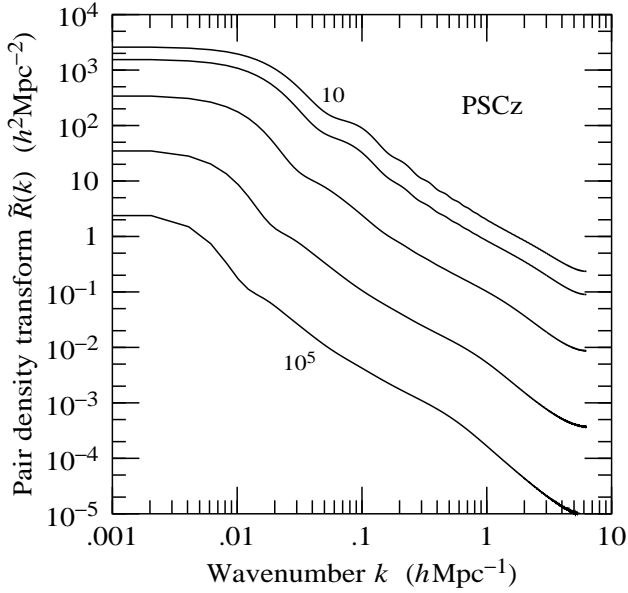


Figure 2. One-dimensional Fourier transform $\tilde{R}(k)$, equation (26), of the pair density $R(r)$, Figure 1, in the PSCz survey. Curves, from top to bottom, are for prior powers $\xi(k_0) = 10, 10^2, 10^3, 10^4$, and $10^5 h^{-3} \text{Mpc}^3$.

random pairs of galaxies in the survey per unit volume interval $4\pi r^2 dr$ at separation r . Although the pair density is commonly computed by Monte Carlo integration, I find it faster, more accurate, and more convenient (since the program was already written) to compute the integral directly, using the procedures described by Hamilton (1993). Figure 1 illustrates the pair density $R(r)$ in the PSCz survey for various values of the prior power $\xi(k_0)$ upon which the survey window $U(r)$ depends, equation (21).

Transforming the Fisher matrix (23) from real space back into Fourier space, $T(k, k') = \int T(r, r') e^{i\mathbf{k} \cdot \mathbf{r} + i\mathbf{k}' \cdot \mathbf{r}'} d^3r d^3r'$, yields

$$T(k, k') = \frac{\pi}{2kk'} [\tilde{R}(k - k') - \tilde{R}(k + k')] \quad (25)$$

where $\tilde{R}(k)$ is the 1-dimensional Fourier transform of the pair density $R(r)$

$$\tilde{R}(k) \equiv 2 \int_0^\infty R(r) \cos(kr) dr. \quad (26)$$

Figure 2 illustrates the pair density transform $\tilde{R}(k)$ in the PSCz survey. As expected, the pair density transform is peaked to small wavenumber k .

The resolution in k -space with which the power spectrum can be measured is roughly set by the inverse scale of the survey, so it makes sense to discretize the Fisher matrix $T(k_m, k_n)$ on a uniform grid in k . A uniform grid in k is also a natural choice numerically, both in evaluating the pair density transform $\tilde{R}(k)$, equation (26), by fast Fourier transform, and then in constructing the Fisher matrix (25) from $\tilde{R}(k)$ evaluated at $k = k_m \pm k_n$.

I proceed numerically as follows. Firstly, compute the pair density $R(r)$, using the method detailed by Hamilton (1993), over a table of separations r and constants $\xi(k_0)$, log-

arithmically spaced at 0.02 dex in r , and 0.1 dex in $\xi(k_0)$. To facilitate taking a fast Fourier transform, interpolate on $R(r)$ to define a new table linearly spaced typically at intervals $0.5 h^{-1} \text{Mpc}$ in r . Optionally zero-pad this table, typically to a separation of $2048 h^{-1} \text{Mpc}$, which results in a grid spacing in k -space of $\Delta k = \pi / (2048 h^{-1} \text{Mpc}) = 0.00153 h \text{Mpc}^{-1}$. Although zero-padding increases the resolution with which the Fisher matrix and its eigenfunctions are sampled, it does not increase the physical resolution available from the survey. For each matrix element $T(k_m, k_n)$, choose $k_0 = (k_m + k_n)/2$, equation (22); set $\xi(k_0)$ to be the $\Omega = 1$ power spectrum proposed by Peacock (1997); interpolate logarithmically on the table of the pair density $R(r)$ as a function of $\xi(k_0)$ to find $R(r)$ for the desired $\xi(k_0)$; fast Fourier (cosine) transform $R(r)$ to find $\tilde{R}(k)$, equation (26); and thence obtain $T(k_m, k_n)$, equation (25). Since $k_0 = (k_m + k_n)/2$ is constant along cross-diagonals of the Fisher matrix $T(k_m, k_n)$, the fast Fourier transform need be carried out only for each cross-diagonal, not for each element of the matrix.

The above procedure yields the Fisher matrix $T(k_m, k_n)$ on a uniformly spaced grid of $k_n = n\Delta k$ with $n = 0, 1, \dots, N$. According to §2, bases of kernels yielding statistically orthogonal sets of windowed power spectra are constructed by diagonalizing matrices of the form $\gamma T \gamma$, where γ is some strictly positive definite symmetric matrix. In all cases considered in the present paper the scaling matrix γ is diagonal in Fourier space, that is, $\gamma(k)$ is some strictly positive function of k . The equation to be solved is then the eigenvalue equation (19). In discretized form, this eigenvalue equation (19) becomes (there is no implicit summation in the following equation, nor elsewhere in this paper)

$$\sum_{n=0}^N \gamma_m F_{mn} \gamma_n u_{n\alpha} = \mu_\alpha u_{m\alpha} \quad (27)$$

where $\gamma_m \equiv \gamma(k_m)$, the matrix F_{mn} is a discrete version of the Fisher matrix T

$$F_{mn} \equiv k_m T(k_m, k_n) k_n 4\pi \Delta k / (2\pi)^3, \quad (28)$$

the eigenvalues μ_α are the same as in equation (19), and the eigenfunctions $u_{m\alpha} \equiv u_\alpha(k_m)$, satisfying the orthonormality condition $\sum_{m=0}^N u_\alpha(k_m) u_\beta(k_m) = \delta_{\alpha\beta}$, are discrete versions of the eigenfunctions $\phi_\alpha(k_m)$

$$u_{m\alpha} \equiv u_\alpha(k_m) \equiv k_m \phi_\alpha(k_m) [4\pi \Delta k / (2\pi)^3]^{1/2}. \quad (29)$$

It is useful also to define corresponding discrete versions $v_{m\alpha} \equiv v_\alpha(k_m)$ of the kernels $\psi_\alpha(k_m)$

$$v_{m\alpha} \equiv v_\alpha(k_m) \equiv k_m \psi_\alpha(k_m) [4\pi \Delta k / (2\pi)^3]^{1/2} \quad (30)$$

which are related to the discrete eigenfunctions $u_\alpha(k_m)$ by the same formula (16) which relates the kernels $\psi_\alpha(k_m)$ and the eigenfunctions $\phi_\alpha(k_m)$, with the same normalization constants c_α ,

$$u_\alpha(k_m) = \gamma(k_m) c_\alpha v_\alpha(k_m). \quad (31)$$

If the kernels $\psi_\alpha(k)$ are normalized on the quadratic convention (4), then the discrete kernels $v_\alpha(k_m)$ are similarly quadratically normalized, $\sum_{m=0}^N v_\alpha(k_m)^2 = 1$. In terms of the discrete kernel $v_\alpha(k_m)$, the windowed power spectrum ξ_α , equation (1), is

$$\xi_\alpha = \sum_{m=0}^N v_\alpha(k_m) \xi(k_m) k_m [4\pi\Delta k/(2\pi)^3]^{1/2}. \quad (32)$$

Although the indices m and n in the eigenvalue equation (27) run over values 0 to N , it follows from equation (25), and the periodicity of the discrete cosine transform $\hat{R}(k_n)$ over the interval $n = [-N, N]$, that the perimeter rows and columns of F are identically zero, $F_{mn} = 0$ for m or $n = 0$ or N . Amongst other things, this guarantees that each eigenfunction $u_\alpha(k_m)$, and consequently each kernel $v_\alpha(k_m)$, vanishes at the boundary values $m = 0$ and N . Thus F_{mn} is effectively reduced to a $(N-1) \times (N-1)$ matrix, with m and n running over values 1 to $N-1$.

Solution of equation (27) yields $N-1$ orthonormal eigenvectors $u_\alpha(k_m)$, and the kernels $v_\alpha(k_m)$ then follow from equation (31). The eigenfunctions $\phi_\alpha(k_m)$ and the corresponding kernels $\psi_\alpha(k_m)$ are given in terms of their discrete counterparts $u_\alpha(k_m)$ and $v_\alpha(k_m)$ by equations (29) and (30). The expected variance $\langle \Delta \hat{\xi}_\alpha^2 \rangle$ of estimates $\hat{\xi}_\alpha$ of the power spectrum windowed through the kernel ψ_α , equations (1) and (2), is related to the eigenvalue μ_α of $\gamma F \gamma$ in equation (27) by equation (10).

In practice, I do not attempt to solve for the eigenfunctions of the full matrix $\gamma F \gamma$, but truncate the matrix typically at a half-wavelength of $\pi/(N\Delta k) = 8 h^{-1} \text{Mpc}$. At a resolution of $\Delta k = \pi/(2048 h^{-1} \text{Mpc})$, this means truncating the matrix to the first $N-1 = 255$ rows and columns (i.e., I only bother to calculate these elements). The cutoff at $8 h^{-1} \text{Mpc}$ seems reasonable, given that the Fisher matrix calculated here assumes Gaussian fluctuations, which may be valid on linear scales, down to $\sim 8 h^{-1} \text{Mpc}$, but certainly fails below this.

4 SOME TRIAL KERNELS

As discussed in §2, there are many distinct bases of kernels $\psi_\alpha(k)$ which yield statistically orthogonal sets of windowed power spectra $\hat{\xi}_\alpha$. Any such basis can be constructed by choosing a strictly positive definite scaling matrix γ , and finding the eigenfunctions $\phi_\alpha = \gamma c_\alpha \psi_\alpha$ of the scaled Fisher matrix $\gamma T \gamma$. In all the cases considered here, the scaling matrix γ is taken to be diagonal in Fourier space, so that it becomes a scaling function $\gamma(k)$.

In this section I show, for the PSCz geometry, bases of kernels for two natural choices of the scaling function, $\gamma(k) = 1$ in §4.1, and $\gamma(k) = \xi(k)$ in §4.2.

As will be seen, the resulting kernels are too broad and wiggly in k -space to be of much practical use. The difficulty is resolved in the next Section, §5, where a basis of kernels which are narrow and positive in Fourier space will be constructed.

4.1 Eigenfunctions of $\langle \Delta \hat{\xi}(k) \Delta \hat{\xi}(k') \rangle$

A natural choice of the scaling function $\gamma(k)$ is a constant $\gamma(k) = 1$. (33)

The kernels $\psi_\alpha(k)$ in this case coincide with the scaled kernels $\phi_\alpha(k)$, and are the eigenfunctions of the covariance matrix $\langle \Delta \hat{\xi}(k) \Delta \hat{\xi}(k') \rangle$. As discussed in §3, these eigenfunctions are in practice constructed as eigenfunctions of the Fisher

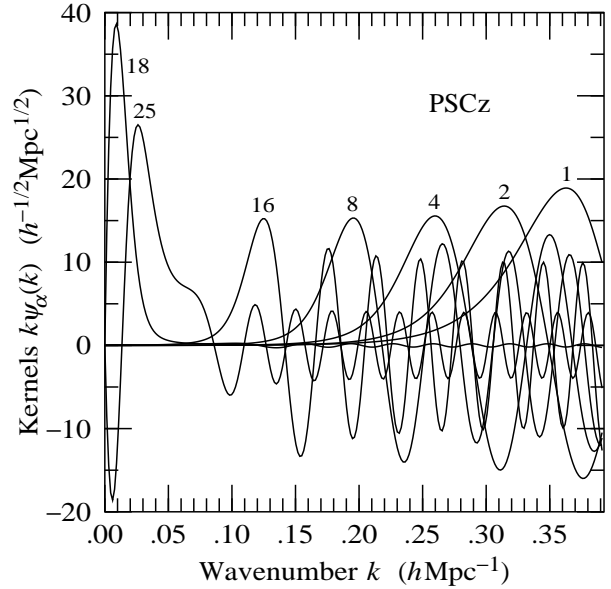


Figure 3. A sample of eigenfunctions $\psi_\alpha(k)$ of the expected covariance $\langle \Delta \hat{\xi}(k) \Delta \hat{\xi}(k') \rangle$ of minimum variance estimates of the power spectrum of the PSCz survey. The eigenfunctions are labelled in order of their eigenvalues, the expected variances $\langle \Delta \hat{\xi}_\alpha^2 \rangle$, the lowest numbered eigenfunction having the smallest variance. The grid in k here is $N = 256$ long, so there are $N-1 = 255$ eigenmodes. The resolution is $\Delta k = \pi/(2048 h^{-1} \text{Mpc}) = 0.00153 h \text{Mpc}^{-1}$.

matrix $T(k, k') = \langle \Delta \hat{\xi}(k) \Delta \hat{\xi}(k') \rangle^{-1}$. The kernels $\psi_\alpha(k)$ here form an orthonormal set, a property unique to the case of a constant scaling function.

Figure 3 shows a selection of eigenfunctions $\psi_\alpha(k)$ of the Fisher matrix $T(k, k')$ of the PSCz survey. The Figure shows $k\psi_\alpha(k)$ rather than $\psi_\alpha(k)$ in part because the discretized version $v_\alpha(k)$ of the kernel, equation (30), includes the extra factor of k , and in part because $k\psi_\alpha(k)$ gives a better impression of the contribution to the windowed power spectrum ξ_α from the power $\xi(k)$ at different k , given that the kernel $\psi_\alpha(k)$ gets multiplied by $k^2 \xi(k)$ and that $\xi(k) \propto k^{-1}$ is a reasonable intermediate choice for the slope of the power spectrum.

The kernels shown in Figure 3 have the drawbacks that for the most part they are neither everywhere positive, nor compact in k . A third problem is that when the range in k is extended to shorter wavelengths (larger k), the fundamental mode at the shortest scales retreats to the smallest available scale. These defects are redressed in §5.

The kernels shown in Figure 3 exhibit some features which are typical also for other choices of the scaling function $\gamma(k)$. The first several (here about 30) modes cover the full range of k with relatively broad kernels of relatively low variance. Subsequent modes (not shown) retrace the same range with ever more oscillatory kernels. The shapes of the noisiest kernels resemble violin notes viewed on an oscilloscope. Unlike the fundamental at small scales (large k), the fundamental at large scales (small k) appears robust against changes of resolution in either real or Fourier space, as long as the resolution in Fourier space is high enough.

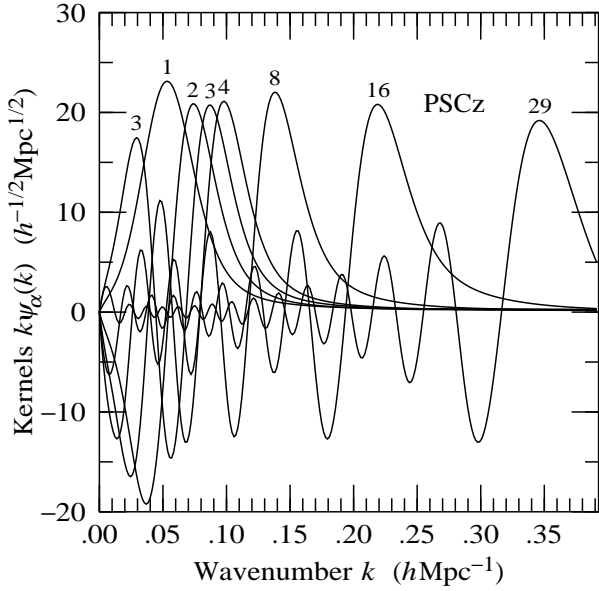


Figure 4. A sample of kernels $\psi_\alpha(k) \propto \phi_\alpha(k)/\xi(k)$ obtained from the eigenfunctions $\phi_\alpha(k)$ of the expected relative covariance $\xi(k)^{-1}\langle\Delta\hat{\xi}(k)\Delta\hat{\xi}(k')\rangle\xi(k')^{-1}$ of minimum variance estimates of the power spectrum of the PSCz survey. The eigenfunctions are labelled in order of the eigenvalues, the relative variances $\langle(\Delta\hat{\xi}/\xi)_\alpha^2\rangle$, equation (38), the lowest eigenfunction having the smallest relative variance. The fundamental mode is centred about the peak in the prior power spectrum at $k \approx 0.05 \text{ hMpc}^{-1}$.

The large scale fundamental, which is almost but not quite positive everywhere, happens to have the 18th smallest variance $\langle\Delta\hat{\xi}_\alpha^2\rangle$. One should bear in mind that the variance depends on the normalization of the kernel, which is taken here to be quadratic, equation (4). A linear normalization would tend to increase the amplitude and hence variance of the wigglier kernels, and would improve the rating of the fundamental. However, the ordering of the variances is correct for the kernels as normalized in Figure 3.

4.2 Eigenfunctions of $\xi(k)^{-1}\langle\Delta\hat{\xi}(k)\Delta\hat{\xi}(k')\rangle\xi(k')^{-1}$

For a contiguous survey of extremely large (tending to infinity) volume V , and for Gaussian fluctuations, the power spectrum itself becomes statistically orthogonal, with variance (Paper 1, eq. [65] with [63])

$$\langle\Delta\hat{\xi}(k)\Delta\hat{\xi}(k')\rangle = (2\pi)^3\delta_D(k-k')\frac{2\xi(k)^2}{V} \quad (34)$$

where the delta-function $(2\pi)^3\delta_D(k-k')$ is the unit matrix in the Fourier representation. It follows from equation (34) that the relative covariance $\xi^{-1}\langle\Delta\hat{\xi}\Delta\hat{\xi}\rangle\xi^{-1}$ of the power spectrum in an extremely large survey is proportional to the unit matrix

$$\frac{\langle\Delta\hat{\xi}(k)\Delta\hat{\xi}(k')\rangle}{\xi(k)\xi(k')} = (2\pi)^3\delta_D(k-k')\frac{2}{V}. \quad (35)$$

This suggests that it might be interesting to consider kernels $\psi_\alpha(k)$ derived from the eigenfunctions $\phi_\alpha(k)$ of the

relative covariance matrix $\xi(k)^{-1}\langle\Delta\hat{\xi}(k)\Delta\hat{\xi}(k')\rangle\xi(k')^{-1}$, corresponding to the case where the scaling function $\gamma(k)$ is chosen equal to the prior power spectrum $\xi(k)$

$$\gamma(k) = \xi(k). \quad (36)$$

The eigenvalues λ_α of the relative covariance matrix $\xi(k)^{-1}\langle\Delta\hat{\xi}(k)\Delta\hat{\xi}(k')\rangle\xi(k')^{-1}$ are equal to the relative variances $\langle(\Delta\hat{\xi}/\xi)_\alpha^2\rangle$

$$\lambda_\alpha = \langle(\Delta\hat{\xi}/\xi)_\alpha^2\rangle \quad (37)$$

defined by

$$\langle(\Delta\hat{\xi}/\xi)_\alpha^2\rangle \equiv \frac{\langle\Delta\hat{\xi}_\alpha^2\rangle}{\int_0^\infty [\psi_\alpha(k)\xi(k)]^2 4\pi k^2 dk / (2\pi)^3}. \quad (38)$$

The relative variances $\langle(\Delta\hat{\xi}/\xi)_\alpha^2\rangle$, which can be regarded as being defined for arbitrary kernels $\psi_\alpha(k)$ by equation (38), have units of inverse volume, and are equal to $2/V$ in the limit of a very large volume, according to equation (35). It may plausibly be argued that the relative variances $\langle(\Delta\hat{\xi}/\xi)_\alpha^2\rangle$ may offer a better measure of noisiness than the absolute variances $\langle\Delta\hat{\xi}_\alpha^2\rangle$, since equation (34) shows that the variance of the power spectrum windowed through some kernel $\psi_\alpha(k)$ may be large not because of noise, but simply because $\xi(k)$ itself is large over the range of the kernel $\psi_\alpha(k)$.

Figure 4 shows the resulting kernels $\psi_\alpha(k) \propto \phi_\alpha(k)/\xi(k)$ derived from the eigenfunctions $\phi_\alpha(k)$ of the relative covariance matrix $\xi(k)^{-1}\langle\Delta\hat{\xi}(k)\Delta\hat{\xi}(k')\rangle\xi(k')^{-1}$. As usual, the eigenfunctions are in practice found as eigenfunctions of the inverse of this matrix, the scaled Fisher matrix $\xi(k)T(k,k')\xi(k')$.

The kernels are labelled not in order of the absolute variances $\langle\Delta\hat{\xi}_\alpha^2\rangle$ as in Figure 3, but rather in order of the eigenvalues, the relative variances $\langle(\Delta\hat{\xi}/\xi)_\alpha^2\rangle$, equation (38). The single fundamental mode has the smallest relative variance, and is centred around the maximum at $\sim 0.05 \text{ hMpc}^{-1}$ in the prior power spectrum.

The kernels shown in Figure 4 share the same problem as those in Figure 3, that they are neither compact nor everywhere positive in Fourier space.

5 A BASIS OF POSITIVE COMPACT KERNELS

It is evidently desirable that a basis of kernels, besides yielding a statistically orthogonal set of windowed power spectra, should be positive and compact in Fourier space. Positivity would preserve the intrinsic positivity of the power spectrum, while compactness would allow the power spectrum to be measured with high resolution in Fourier space. In this Section I show how to construct such a basis. A numerically convenient version of the construction is described in the paragraph containing equations (45) and (46).

To understand how to make kernels compact and positive, consider the mathematical theorem that the eigenmodes of a real symmetric matrix are unique up to arbitrary orthogonal rotations amongst degenerate eigenmodes, that is, amongst eigenmodes with the same eigenvalue. In the present case the eigenvalues of the Fisher matrix, even if not degenerate, are nevertheless many and finely spaced, and

there is much potential for nearly degenerate eigenmodes to mix in random unpleasant ways. This suggests that mixing can be reduced in Fourier space by lifting the degeneracy of eigenvalues in Fourier space, which can be accomplished by applying a scaling function $\gamma(k)$ which is a strongly varying function of k . Numerical experiment indicates that the most compact and positive kernels arise from choosing the scaling function $\gamma(k)$ to be as steep as possible.

The limiting case of an infinitely steep scaling function $\gamma(k)$ can be solved analytically, and as will be seen, Figure 5, it gives rise to a basis of positive kernels as pretty as one could wish for. The solution, given immediately following, is simpler than for finite scaling functions $\gamma(k)$, requiring no diagonalization at all.

Suppose that the matrix $T(k_m, k_n)$ is discretized, as is the practical case, with $m, n = 1, 2, \dots, N-1$, and suppose that $\gamma(k_1) \gg \gamma(k_2) \gg \dots \gg \gamma(k_{N-1})$. For a uniformly spaced grid of k , as here, it is convenient to work with the discrete version F_{mn} , equation (28), of the Fisher matrix, and the associated discrete versions $u_\alpha(k_m)$ and $v_\alpha(k_m)$ of the eigenfunctions $\phi_\alpha(k_m)$ and the kernels $\psi_\alpha(k_m)$, equations (29) and (30). The equation to be solved is the discrete eigenvalue equation (27).

Since the first column of the scaled Fisher matrix $\gamma_m F_{mn} \gamma_n$ (no implicit summation) is huge compared to all other columns, the first eigenfunction $u_1(k_m)$, the one with largest eigenvalue, is proportional to this first column, $u_1(k_m) \propto \gamma_m F_{m1}$. It follows that the first kernel $v_1(k_m) \propto u_1(k_m)/\gamma_m$ is proportional to the first column of the Fisher matrix, $v_1(k_m) \propto F_{m1}$. Since the first element of the eigenvector $u_1(k_m)$ is huge compared to all other elements, the eigenvector is $u_1(k_m) = (1, 0, \dots, 0)$ in the limit, but the higher elements of the kernel vector $v_1(k_m)$ remain generally nonzero.

The remaining eigenvectors $u_\alpha(k_m)$, $\alpha = 2, \dots, N-1$, must be orthogonal to the first eigenvector $u_1(k_m)$, which means that their first elements must be zero, $u_\alpha(k_1) = 0$ for $\alpha > 1$. Since the first two columns of the scaled Fisher matrix are huge compared to the others, the second eigenfunction $u_2(k_m)$, with the second largest eigenvalue, must be some linear combination of the first two columns $\gamma_m F_{m1}$ and $\gamma_m F_{m2}$, and it must be proportional to that linear combination whose first element vanishes, $u_2(k_m) \propto \gamma_m (F_{11}F_{m2} - F_{12}F_{m1})$. The second kernel follows from $v_2(k_m) \propto u_2(k_m)/\gamma_m \propto F_{11}F_{m2} - F_{12}F_{m1}$, whose first element is zero but whose other elements are finite. The second eigenfunction is $u_2(k_m) = (0, 1, 0, \dots, 0)$ in the limit. Proceeding, one concludes that the α th eigenfunction $u_\alpha(k_m)$ is that linear combination of the first α columns of the scaled Fisher matrix which is zero in the first $\alpha-1$ elements. Normalized to $u_\alpha(k_\alpha) = 1$, the eigenfunction is

$$u_\alpha(k_m) = \begin{cases} \frac{\gamma_\alpha F_{m\alpha}^{-1}}{\gamma_m F_{\alpha\alpha}^{-1}} & (m \leq \alpha) \\ \frac{\gamma_m}{\gamma_\alpha} \sum_{n=1}^{\alpha-1} F_{mn} F_{n\alpha}^{-1} & (m \geq \alpha) \end{cases} \quad (39)$$

where $F_{m\alpha}^{-1}$ denotes the (m, α) th element of the inverse of the $\alpha \times \alpha$ submatrix consisting of the first α rows and columns of F_{mn} . In the limit, the eigenfunction $u_\alpha(k_m)$ is 1 in the α th place and zero in all others:

$$u_\alpha(k_m) \rightarrow \begin{cases} 1 & (m = \alpha) \\ 0 & (m \neq \alpha) \end{cases}. \quad (40)$$

The associated eigenvalue of the scaled Fisher matrix is

$$\mu_\alpha = \frac{\gamma_\alpha^2}{F_{\alpha\alpha}^{-1}}. \quad (41)$$

It follows from equation (39) that the kernel $v_\alpha(k_m) \propto u_\alpha(k_m)/\gamma_m$ is given by

$$v_\alpha(k_m) = c'_\alpha{}^{-1} \sum_{n=1}^{\alpha} F_{mn} F_{n\alpha}^{-1} \quad (42)$$

where c'_α is an arbitrary normalization constant. Note $v_\alpha(k_m) = 0$ for $m \leq \alpha-1$. For the quadratic normalization convention adopted in this paper, $\sum_{m=1}^{N-1} v_\alpha(k_m)^2 = 1$, the normalization constant is $c'_\alpha = \left[\sum_{m=1}^{N-1} \left(\sum_{n=1}^{\alpha} F_{mn} F_{n\alpha}^{-1} \right)^2 \right]^{1/2}$. The usual normalization constants c_α which relate ϕ_α and ψ_α , equation (16), or equivalently u_α and v_α , equation (31), are related to the normalization constants c'_α above by

$$c_\alpha = c'_\alpha / \gamma_\alpha. \quad (43)$$

The estimators $\hat{\xi}_\alpha$, equations (1), (2), and (32), of power spectra windowed through the kernels $\psi_\alpha(k)$ corresponding to $v_\alpha(k)$, equation (30), are by construction statistically orthogonal, in accordance with equation (7). The expected variance $\langle \Delta \hat{\xi}_\alpha^2 \rangle$ is given by formula (10), which here reduces to

$$\langle \Delta \hat{\xi}_\alpha^2 \rangle = \frac{F_{\alpha\alpha}^{-1}}{c'^2_\alpha} \quad (44)$$

the factors of γ_α in the eigenvalue μ_α , equation (41), and the normalization constant c_α , equation (43), cancelling out.

In practice, the kernels $v_\alpha(k_m)$, equation (42), can be evaluated conveniently by the following procedure, which reduces the Fisher matrix F_{mn} iteratively to lower triangular form. Suppose that the discrete Fisher matrix F_{mn} defined by equation (28) has been reduced so far to the point where all the above diagonal elements of the first $\alpha-1$ rows are zero. Call this reduced Fisher matrix F'_{mn} , which starts equal to the Fisher matrix F_{mn} at step one, $\alpha = 1$. Then the α th kernel $v_\alpha(k_m)$ is proportional to the α th column of the reduced Fisher matrix

$$v_\alpha(k_m) = \frac{F'_{m\alpha}}{c''_\alpha} \quad (45)$$

where c''_α is an arbitrary normalization constant. For the quadratic normalization convention $\sum_{m=1}^{N-1} v_\alpha(k_m)^2 = 1$, the normalization constant is $c''_\alpha = \left(\sum_{m=1}^{N-1} F'^2_{m\alpha} \right)^{1/2}$. The previous normalization constant c'_α in equation (42) is related to the normalization constant c''_α by $c'_\alpha = c''_\alpha F_{\alpha\alpha}^{-1}$. The quantity $F_{\alpha\alpha}^{-1}$, which was defined immediately following equation (39), and upon which the eigenvalue μ_α , equation (41), also depends, is equal to the reciprocal of the diagonal element of the reduced Fisher matrix, $F_{\alpha\alpha}^{-1} = 1/F'_{\alpha\alpha}$. Thus the eigenvalue (41) is $\mu_\alpha = \gamma_\alpha^2 F'_{\alpha\alpha}$. The variance $\langle \Delta \hat{\xi}_\alpha^2 \rangle$ of the windowed power spectrum $\hat{\xi}_\alpha$ is

$$\langle \Delta \hat{\xi}_\alpha^2 \rangle = \frac{F'_{\alpha\alpha}}{c''^2_\alpha}. \quad (46)$$

Now zero out the α th row of F'_{mn} above the diagonal

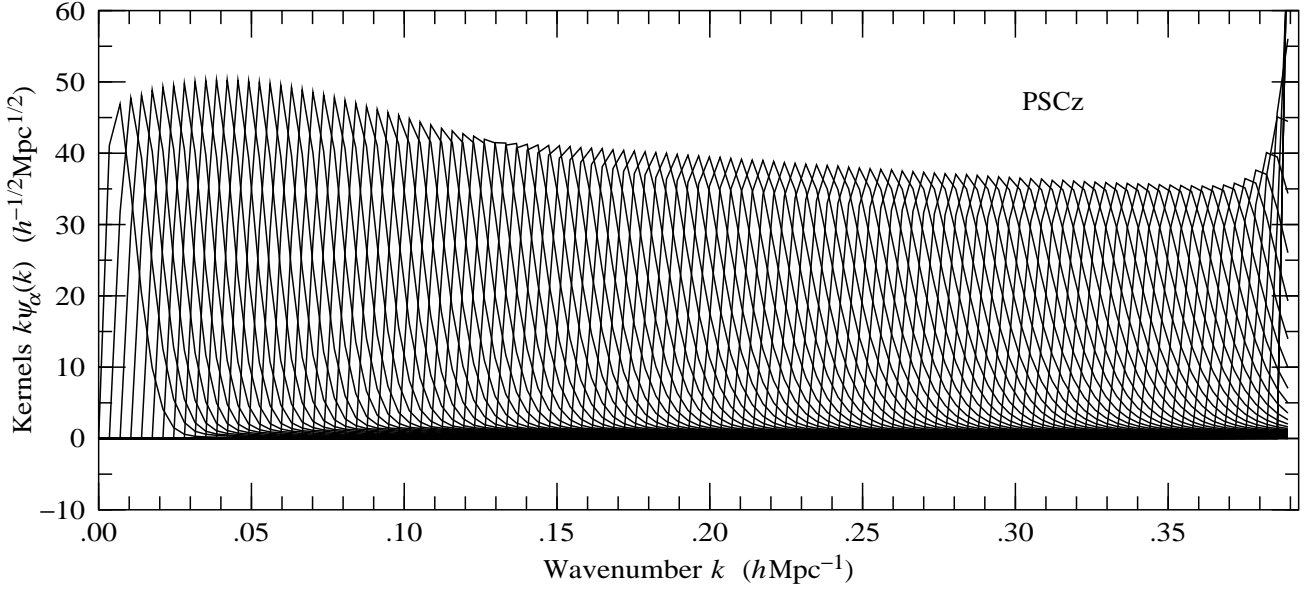


Figure 5. Basis of positive kernels obtained by applying an infinitely steep scaling function to the Fisher matrix of the PSCz survey, as described in the text, §5. Counterintuitively, notwithstanding the overlap between kernels, the minimum variance estimators $\hat{\xi}_\alpha$ of power spectra windowed through these kernels are statistically orthogonal. The resolution is $\Delta k = \pi/(896 h^{-1} \text{Mpc}) = 0.00351 h \text{Mpc}^{-1}$, which is the highest resolution such that all the eigenvalues are positive. The grid in k is $N = 112$ long, so there are $N - 1 = 111$ kernels. The crowding of the kernels at the right of the graph is an artefact of truncating the Fisher matrix at $k_N = N\Delta k = \pi/(8 h^{-1} \text{Mpc}) = 0.393 h \text{Mpc}^{-1}$, and could easily be avoided by extending the matrix to shorter wavelengths (larger k).

by subtracting an appropriate multiple of the α th column $F'_{m\alpha}$ from each higher column, that is, $F'_{mn} \rightarrow F'_{mn} - F'_{m\alpha} F'_{\alpha n} / F'_{\alpha\alpha}$ for $n \geq \alpha + 1$. Relabel the resulting matrix again F'_{mn} . Repeat the above steps until the matrix is in lower triangular form.

When done, the columns of the lower triangular matrix $F'_{m\alpha}$ will be proportional to the kernels $v_\alpha(k_m)$, equation (45), while the diagonal elements $F'_{\alpha\alpha}$ will be proportional to the variances $\langle \Delta \hat{\xi}_\alpha^2 \rangle$, equation (46). Note that the first $\alpha - 1$ elements of the α th kernel are zero, $v_\alpha(k_m) = 0$ for $m \leq \alpha - 1$; in particular, the final kernel is $v_{N-1}(k_m) = (0, \dots, 0, 1)$.

If the Fisher matrix were strictly positive definite, that is, if all its eigenvalues were strictly positive, as should be the case theoretically, then all the variances $\langle \Delta \hat{\xi}_\alpha^2 \rangle$, equation (46), would be guaranteed to be positive. As remarked near the beginning of §3, the eigenvalues of the approximate Fisher matrix of the PSCz survey computed here are all positive only if the resolution is taken to be $\Delta k = \pi/(896 h^{-1} \text{Mpc}) = 0.00351 h \text{Mpc}^{-1}$ or coarser. This, apparently, sets the natural resolution of the PSCz survey.

Figure 5 shows the resulting basis of kernels for the PSCz survey at the ‘natural’ resolution $\Delta k = \pi/(896 h^{-1} \text{Mpc})$. All of the $N - 1 = 111$ eigenvalues down to the cutoff half-wavelength of $\pi/(N\Delta k) = 8 h^{-1} \text{Mpc}$ are positive, and all the 111 kernels are everywhere positive. At higher resolutions an increasing fraction of the eigenvalues of the Fisher matrix are negative. At a resolution of $\Delta k = \pi/(1024 h^{-1} \text{Mpc}) = 0.00307 h \text{Mpc}^{-1}$ there are 3 negative of 127 eigenvalues, while at $\Delta k = \pi/(1536 h^{-1} \text{Mpc}) = 0.00204 h \text{Mpc}^{-1}$ there are 27 negative of 191 eigenvalues. The kernels corresponding to negative eigenvalues are not

positive everywhere, and they are not well-behaved. Kernels with positive eigenvalues remain positive and mostly well-behaved, even when kernels with negative eigenvalues are present, although there is an increasing tendency for kernels, especially those adjacent to kernels with negative eigenvalues, to become shaky at finer resolutions.

It is evident that the kernels in Figure 5 overlap substantially. Yet they yield, by construction, a statistically orthogonal set of windowed power spectra. This seems astonishing: how can it be? The answer is that the non-diagonal elements of the covariance $\langle \Delta \hat{\xi}(k) \Delta \hat{\xi}(k') \rangle$ of the power spectrum of the survey are not all positive, and the positive and negative covariances cancel precisely between overlapping kernels. Rather, it is the elements of its inverse, the Fisher matrix $T(k, k') = \langle \Delta \hat{\xi}(k) \Delta \hat{\xi}(k') \rangle^{-1}$, which are all positive in Fourier space, at least in the approximation used here, equation (20). If the elements of the Fisher matrix are all positive, then necessarily the elements of its inverse cannot all be positive. Numerically, I find that the elements of the (discretized) covariance $\langle \Delta \hat{\xi}(k) \Delta \hat{\xi}(k') \rangle$ in the PSCz survey oscillate wildly between positive and negative values. This is not simply a numerical artefact. In real space, the survey correlation function is a sum $\xi(r_{ij}) + \Phi(\mathbf{r}_i)^{-1} \delta_D(\mathbf{r}_{ij})$ of a cosmic term $\xi(r_{ij})$ and a Poisson sampling term $\Phi(\mathbf{r}_i)^{-1} \delta_D(\mathbf{r}_{ij})$. This survey correlation is ill-defined in Fourier space, because the selection function $\Phi(\mathbf{r}_i)$ vanishes outside the survey, so that the Poisson sampling term $\Phi(\mathbf{r}_i)^{-1}$ blows up. It is only the inverse of the survey correlation, which leads to the Fisher matrix, which is well-behaved in Fourier space.

The crowding of the kernels to the right of Figure 5 is an artefact of truncating the Fisher matrix at a half-wavelength of $\pi/(N\Delta k) = 8 h^{-1} \text{Mpc}$, and could easily be

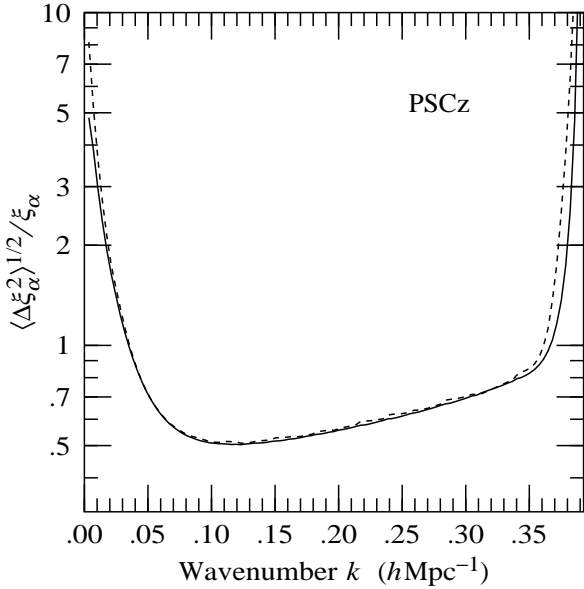


Figure 6. Expected relative uncertainties $\langle \Delta \hat{\xi}_\alpha^2 \rangle^{1/2} / \xi_\alpha$ in the power spectra $\hat{\xi}_\alpha$ windowed through the kernels $\psi_\alpha(k)$ of Figure 5, plotted against $k_\alpha = \alpha \Delta k$. The solid line shows the expected uncertainty in the minimum variance estimator, while the dashed line shows the expected uncertainty in the zeroth order approximation to the minimum variance estimator, as described in Paper 1, equations (97) and (100).

avoided by extending the matrix to shorter wavelengths. I show the crowding to illustrate what happens. The kernels remain compact, positive, and well-behaved, with no negative eigenvalues, if the matrix is extended by a factor of 4 to $N = 448$ at the same resolution, corresponding to a minimum half-wavelength of $\pi/(N\Delta k) = 2 h^{-1}\text{Mpc}$. Presumably the kernels continue to remain well-behaved beyond this.

The kernels in Figure 5 are narrowest and highest at the peak $\sim 0.05 h\text{Mpc}^{-1}$ in the prior power spectrum. The width, hence height, of a kernel in this basis is approximately equal to the width of the pair density transform $\hat{R}(k)$ at the corresponding wavenumber k , which in turn is related to the prior power spectrum as illustrated in Figure 2. Increasing the prior power spectrum $\xi(k)$ would narrow the kernel, while reducing the prior power spectrum would broaden it. Figure 2 shows that, for the PSCz survey, changing the prior power spectrum by a factor of ten would change the width of the kernel by a factor of somewhat less than two. If I had chosen to use as a prior Peacock's (1997) power spectrum for $\Omega = 0.3$, which is somewhat higher than the adopted case $\Omega = 1$, then the kernels in Figure 5 would have been somewhat narrower.

The expected relative uncertainty in the windowed power spectrum $\hat{\xi}_\alpha$ is given by $\langle \Delta \hat{\xi}_\alpha^2 \rangle^{1/2} / \xi_\alpha$. Figure 6 shows this relative uncertainty for the kernels shown in Figure 5. In addition to the uncertainty in the minimum variance estimator, Figure 6 also shows the expected uncertainty in the zeroth order approximation to the minimum variance estimator, equations (97) and (100) of Paper 1. The two uncertainties are quite similar, except for the first few kernels at the smallest wavenumbers k , suggesting that the ze-

roth order approximate estimator should be adequate except at the longest wavelengths. The sharp rise in the uncertainty to the right of the graph (and the discrepancy between the two uncertainties there), is again an avoidable artefact of truncating the Fisher matrix at a half-wavelength of $\pi/(N\Delta k) = 8 h^{-1}\text{Mpc}$. Roughly speaking, the relative variance of a mode α at wavenumber $\sim k_\alpha$ is given by $\langle \Delta \hat{\xi}_\alpha^2 \rangle / \xi_\alpha^2 \approx 2/(VV_k)$, where V is the effective volume of the survey at k_α , the volume over which $\xi(k_\alpha)\Phi(\mathbf{r}) \gtrsim 1$, and $V_k \approx 4\pi k_\alpha^2 \Delta k_\alpha / (2\pi)^3$ is the coherence volume of the mode, Δk_α being the width of the mode. Interestingly, the selection function of the PSCz survey falls off in such a way that the effective volume V initially increases as the wavelength increases (k decreases) somewhat more rapidly than the coherence volume V_k decreases, so causing the relative uncertainty shown in Figure 6 to decline mildly with decreasing k down to $\sim 0.1 h\text{Mpc}^{-1}$.

6 QUESTIONS

This paper leaves unanswered a number of questions which it would be interesting to investigate further.

(1) Notwithstanding the mathematical argument at the beginning of §5 which motivated the idea of applying an infinitely steep scaling function to the Fisher matrix, it remains somewhat mysterious as to why the kernels which result are indeed all positive. Is positivity true for arbitrary survey geometries? Is there some fundamental theorem which ensures positivity?

(2) The kernel set obtained from the infinitely steep scaled Fisher matrix ceased to be positive when the resolution in Fourier space was set too high, and this breakdown was signalled by the appearance of negative eigenvalues in the discretized Fisher matrix. Presumably the appearance of negative eigenvalues is an artefact of the approximation to the Fisher matrix used here, since the eigenvalues of the true Fisher matrix should all be strictly positive. Nevertheless, the numerical results based on the approximate Fisher matrix do make physical sense: somehow the numerics know that the inverse scale of the survey sets a limit to the resolution at which the power spectrum can be probed. It would be nice to have a clearer understanding of what is going on here.

(3) Are there other ways to construct sets of kernels which are strictly positive and yield statistically orthogonal sets of power spectra, besides applying infinitely steep scaling functions? Clearly the kernel set illustrated in Figure 5 is not unique, since there are other ways of ordering the elements of the scaling function than the ordering $\gamma(k_1) \gg \gamma(k_2) \gg \dots \gg \gamma(k_{N-1})$ adopted there. I have checked for example that the inverse ordering $\gamma(k_1) \ll \gamma(k_2) \ll \dots \ll \gamma(k_{N-1})$ also leads to a strictly positive set of kernels for the PSCz survey. Whereas the former ordering leads to kernels which have tails to large k , the opposite ordering leads to kernels with tails to small k .

I conclude with a caveat about applying the results of this paper and Paper 1. The Fisher matrix computed here is for the unredshifted power spectrum, whereas the galaxy positions in 3-dimensional surveys such as the PSCz survey

lie in redshift space, and are subject to distortions from peculiar velocities. Although, as explained in §4 of Paper 1, the Fisher matrix of the unredshifted power spectrum is the same irrespective of whether the data lie in real or redshift space, the minimum variance pair window is not the same in both real and redshift space. While there is no harm in applying a pair window which is not minimum variance (see Paper 1, §7 for further clarification of this point), it would be incorrect to carry over the pair window which is minimum variance in real space and imagine that it is also minimum variance for the power spectrum in redshift space. I hope to address the issue of the optimal measurement of the power spectrum from data in redshift space in a subsequent paper. Such a measurement would include a simultaneous measurement of the linear redshift distortion parameter $\beta \approx \Omega^{0.6}/b$.

7 SUMMARY

This is the second of a pair of papers which address the problem of how to estimate the unredshifted power spectrum of fluctuations from a galaxy survey in optimal fashion.

It has been shown that there are many distinct bases of kernels $\psi_\alpha(k)$ which yield statistically orthogonal sets $\hat{\xi}_\alpha$ of estimates of windowed power spectra. With each distinct basis is associated a positive definite scaling matrix γ . A basis of kernels is constructed by finding the eigenfunctions $\phi_\alpha \propto \gamma\psi_\alpha$ of the scaled Fisher matrix $\gamma T \gamma$.

Illustrative results have been presented for the case of the PSCz survey (Saunders et al., in preparation). The Fisher matrix of the unredshifted power spectrum of this survey was computed assuming Gaussian fluctuations and the zeroth order FKP approximation proposed in Paper 1, together with the prior power spectrum proposed by Peacock (1997). Examples of bases of kernels yielding statistically orthogonal sets of windowed power spectra for this survey have been shown.

The main result of the paper is that, among the many possible bases of kernels which give rise to statistically orthogonal sets of windowed power spectra, there is a particular basis, obtained by applying an infinitely steep scaling function to the Fisher matrix, which leads to kernels which are compact and everywhere positive in Fourier space. The basis is illustrated for the PSCz survey in Figure 5. This basis is easily computed from the Fisher matrix, without even the necessity to diagonalize. The basis of kernels, along with the associated minimum variance pair weighting described in Paper 1, §§2.3 and 7, would appear to offer a solution to the problem of how to measure the unredshifted power spectrum of a galaxy survey in optimal fashion.

ACKNOWLEDGEMENTS

This work was supported by NSF grant AST93-19977 and by NASA Astrophysical Theory Grant NAG 5-2797. I thank Max Tegmark and the referee, Alan Heavens, for helpful comments.

REFERENCES

- Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, 426, 23 (FKP)
- Hamilton A. J. S., 1993, *ApJ*, 417, 19
- Hamilton A. J. S., 1997, *MNRAS*, submitted (Paper 1)
- Lawrence A., Rowan-Robinson M., Crawford J., Parry I., Xia X.-Y., Ellis R. S., Frenk C. S., Saunders W., Efstathiou G., Kaiser N., 1997, for submission to *MNRAS* (QDOT)
- Peacock J. A., 1997, *MNRAS*, submitted
- Tegmark M., Taylor A. N., Heavens A. F., 1997, *ApJ*, submitted
- Turner E. L., 1979, *ApJ*, 231, 645